1. A random sample was selected from Country X’s population to estimate the unemployment rate. The results showed that the 95% confidence interval for the population unemployment rate was between 0.18 and 0.22.

If many samples of the same size were collected by the same procedure, about 95% of these samples will have sample unemployment rate that lie between 0.18 and 0.22.

1. True
2. False

***Explanation:*** *See definition on confidence interval in Chapter 3 Unit 8.*

1. The phrase “95% confidence interval” can be interpreted to mean that:
2. you can be 95% confident that you have selected a sample whose interval does not include the population mean.
3. if all possible populations are taken and confidence intervals are calculated, 95% of those intervals would include the true sample mean somewhere in their interval.
4. if all possible samples are taken and confidence intervals are calculated, 95% of those intervals would include the true sample mean somewhere in their interval.
5. if all possible samples are taken and confidence intervals are calculated, 95% of those intervals would include the true population mean somewhere in their interval.

***Explanation:*** *See definition on confidence interval in Chapter 3 Unit 8.*

1. Two simple random samples were drawn from a population of several million people. However, Sample A has a sample size of 100 and Sample B has a sample size of 1000. The average weights of Sample A and Sample B are used to estimate the average population weight.

Consider the following statements, with respect to estimating the population parameter:

1. Average weight from Sample A is likely to have less random error than the average weight from Sample B.
2. Average weight from Sample A is likely to have less bias than the average weight from Sample B.

Which of the above statements is/are true?

1. Only (I).
2. Only (II).
3. Both (I) and (II).
4. Neither (I) nor (II).

***Explanation****: A sample with larger size is likely to have a smaller random error. See definition on random error in Chapter 3 Unit 7. The mean of a simple random sample is unbiased for the population mean (Chapter 3 Unit 5, slide 64).*

1. Two different samples of size 100 were randomly picked from a population of 5000 durians. For each of the samples, the 95% confidence interval for the average population weight was calculated. What can be said about the average population weight, with respect to the two confidence intervals?
2. It must not lie in any of these confidence intervals.
3. It must lie in exactly one of these confidence intervals.
4. It must lie in both of these confidence intervals.
5. It may or may not lie in any of these confidence intervals.

***Explanation:*** *The confidence intervals may or may not contain the population parameter, which is the average population weight. See definition on confidence interval in Chapter 3 Unit 8, slide 90.*

1. A sampling frame should be \_\_\_\_\_\_ the target population to not lose generalisability of the results from the sample to the population.
   1. equal to
   2. smaller than
   3. larger than
2. (I) only.
3. (II) only.
4. (III) only.
5. (I) or (II).
6. (I) or (III).

***Explanation:*** *Sampling frame chosen should be equal to or larger than the target population to achieve good coverage. See definition on sampling frame in Chapter 3 Unit 1, slide 1.*

1. To find out the employment status of fresh graduates from University ABC, a questionnaire was sent to all of them. 30% of fresh graduates responded to the survey. The employment rate was calculated from the responses. Which of the following can cause this rate to differ from the true rate?
2. Selection bias
3. Non-response bias
4. None of the other options

***Explanation****: Please refer to Chapter 3 Unit 4, slide 43. 30% of fresh graduates may not be representative of the whole population. It is possible that those who did not respond may have different employment status from those who responded. Though the response rate is low, it is still a census. There is no selection bias in a census.*

1. To find out and track the employment status of a community, questionnaires were sent to the residents of this community in the years of 2019 and 2020. Some of them responded to the surveys. The employment rates were calculated from the responses. There are only two blocks in the community. Which of the following may be the reason that the overall employment rate changed more than in blocks A and B?

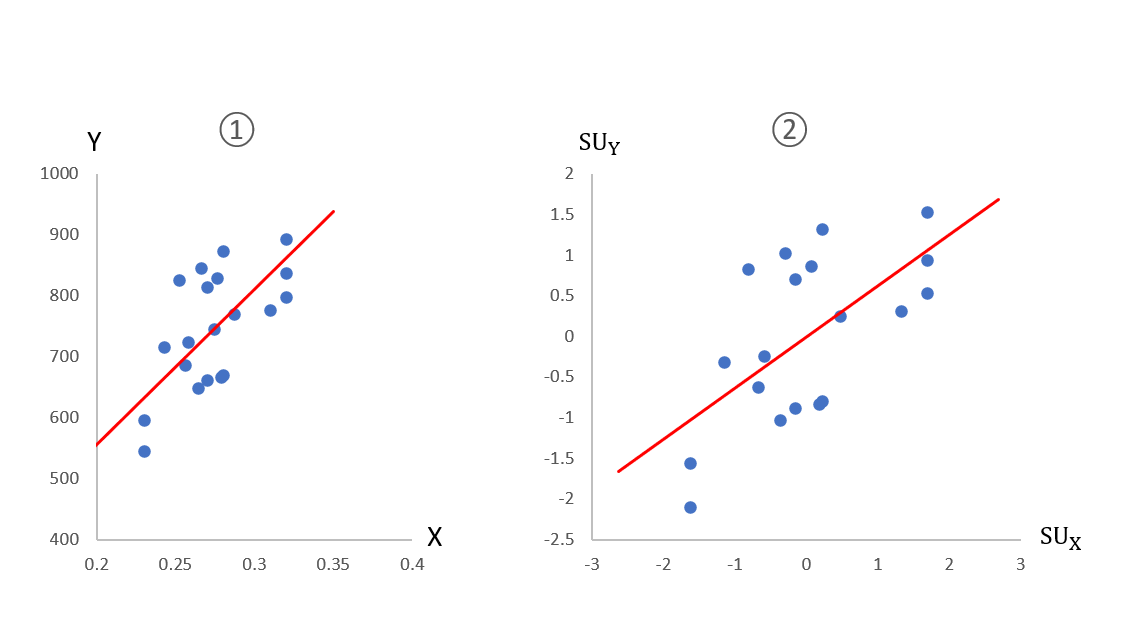
|  |  |  |  |
| --- | --- | --- | --- |
|  | Employment Rate (%) | | |
| 2019 | 2020 | % Change, year-on-year |
| Block A | 100% | 91% | -9 |
| Block B | 91% | 84% | -7 |
| Overall | 98% | 85% | -13 |

1. In 2019, most respondents were from Block A but most respondents were from Block B in 2020.
2. In 2019, most respondents were from Block B but most respondents were from Block A in 2020.
3. Change of overall employment rate should fall between the changes of 2 blocks. The example above is not possible to occur. There might be some calculation errors.

***Explanation****: Please refer to the Annual Fresh Graduate Employment Survey in Chapter 3 Unit 4, slides 51 and 52. Since residents decide whether they wish to take part in the study, the response rate of each block will vary from year to year, which affects the overall rates. In 2019, Block A had more respondents than B, which led to the overall rate being closer to the rate of Block A. In 2020, the opposite occurred, so the overall rate was closer to the rate of Block B.*

1. Let X and Y denote two variables measured on 19 subjects. Let the mean and standard deviation for the X values be written as and ; similarly the mean and SD for the Y values are written as and . After converting the values of X and Y to standard units using the formulas:

the scatter plot changes from ① to ②. The red lines are respectively the regression lines of Y on X, and of on respectively. Which of the following is true?



1. Correlation coefficient is not changed.
2. Correlation coefficient is decreased.
3. Correlation coefficient is increased.

***Explanation****: The change of scale of variables does not affect the correlation coefficient. Please refer to Chapter 2 Unit 4, slide 52.*

1. A telecommunication company asks 100 commuters at a bus station how many mobile phones they own. Which of the following statements is/are true?
2. If among the 100 commuters, every person has 2 or more mobile phones, then the mean number of mobile phones in the sample will be greater than or equal than 2.
3. If the mean number of mobile phones in the sample is greater than or equal than 2, then everyone among the 100 commuters has 2 or more mobile phones.
4. (I) only
5. (II) only
6. Both (I) and (II)
7. *Neither (I) nor (II)*

***Explanation:*** *If all 100 numbers are greater than or equal to 2, the mean will also be greater than or equal to 2. But the opposite is not true: a large mean can result from some people having very large values, while other have low values. For an extreme example, if one person has 101 mobile phones and the rest have only one each, then the mean is 2.*